Fundamental Elements of Computer Based and Computer Adaptive Testing and Update on Key Smarter Balanced Assessment Consortium Activities

Presentation to EDCO Liaisons
February 13, 2013

CALIFORNIA DEPARTMENT OF EDUCATION
Tom Torlakson, State Superintendent of Public Instruction

Fundamental Elements of Computer Based and Computer Adaptive Testing

TOM TORLAKSON
State Superintendent
of Public Instruction

- Smarter Balanced assessments
- Computer based testing
- Computer adaptive testing
- Automated scoring

Smarter Balanced Computer Based Tests

- Summative computer based and adaptive
  - "Adaptivity" enhancement of computer based

- 2013 pilot test: computer based, not adaptive

- 2014 field test: start at a fixed level and becomes adaptive as student progresses through items

Advantages of Computer Based Testing (CBT)

- Can present items with response formats not available with paper and pencil (P&P) tests

- Can increase test security by varying stimuli

- Dramatically reduces chance of correct answer by guessing

- Can provide nearly instant results

- Speed of response can be used as additional info on proficiency (Smarter Balanced will not use this "measurement.")

4

Advantages of CBT (cont.)

- No more lost or damaged answer sheets
- Field testing of items is simplified
- Faulty items can be immediately removed
- Provide more engaging stimuli
- Better measurement of problem solving, critical thinking, and analytic skills
- Dramatically reduced cost of scoring constructed response items

TOM TORLAKSON
State Superintendent
of Public Instruction

Important Issues in CBT

- Administration logistics more complicated, technical support required

- Equipment for testing more costly

- Technology requirements

- Possibility of power/equipment failure interrupting testing

- System functionality, reliability, and recovery more complicated

- Unfamiliarity with computers may disadvantage some students

TOM TORLAKSON
State Superintendent
of Public Instruction

6

CBT Stimulus and Response Formats

- Greater variety of questions

- Animated stimuli possible

- Audio recordings can be used for aural comprehension of spoken language

- Videos or simulations can replace long descriptions of scenarios

- Fill in the blank and short answer questions

- Drag and drop responses

- Students may graph answers

- Students may edit actual documents

Computer Adaptive Testing (CAT)

- Presents items to students that match their ability

- Students take base items to determine starting level

- Get progressively harder/easier items based on performance

- May be based on individual items or groups of items

TOM TORLAKSON
State Superintendent
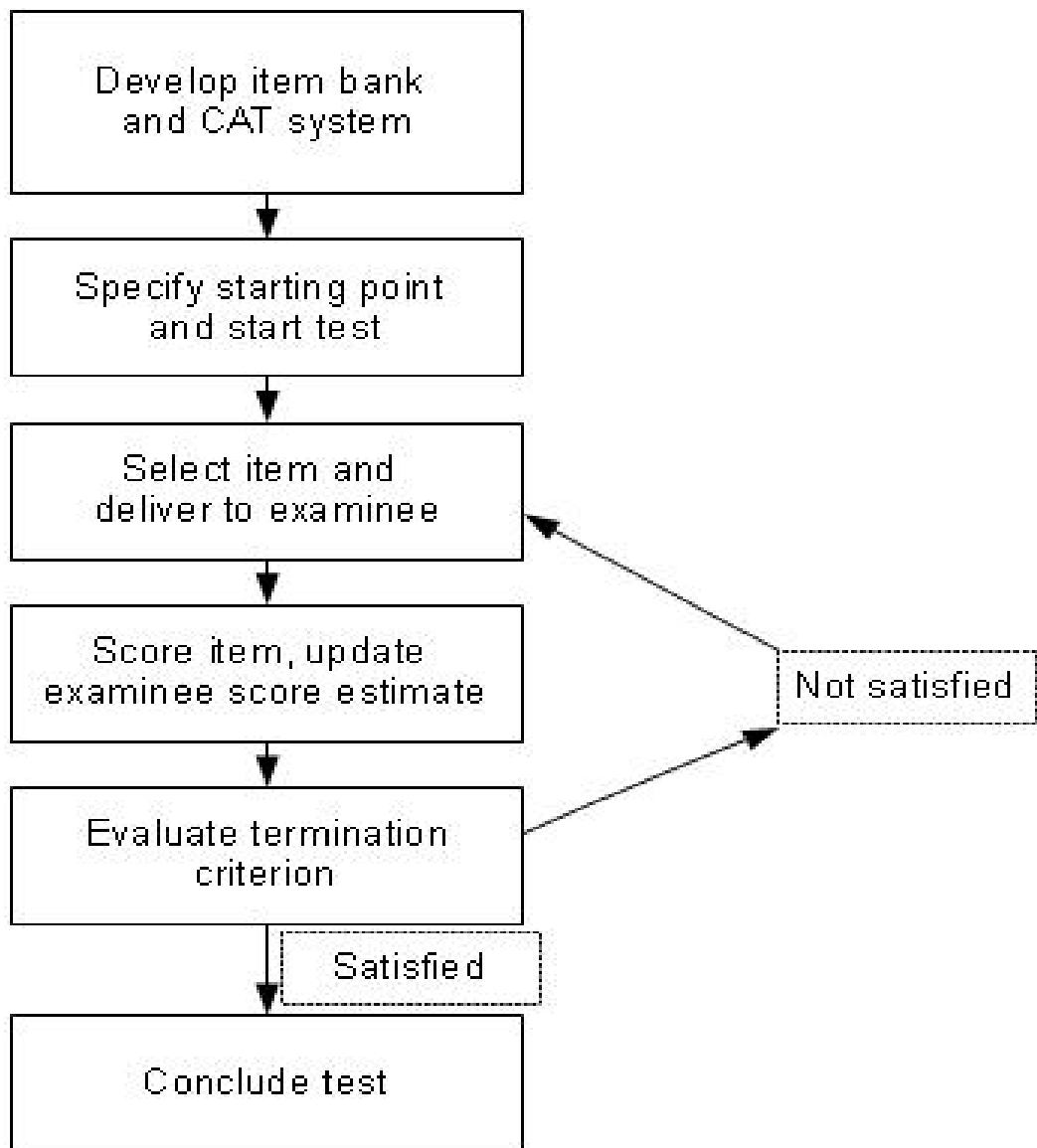of Public Instruction

Advantages of CAT

- Enhanced measurement precision for very low/high performing students

- Can measure all students more accurately with fewer items

- Provides better security than a fixed-form test

- Can be designed to measure growth

- Decreases likelihood of students becoming discouraged or bored

TOM TORLAKSON
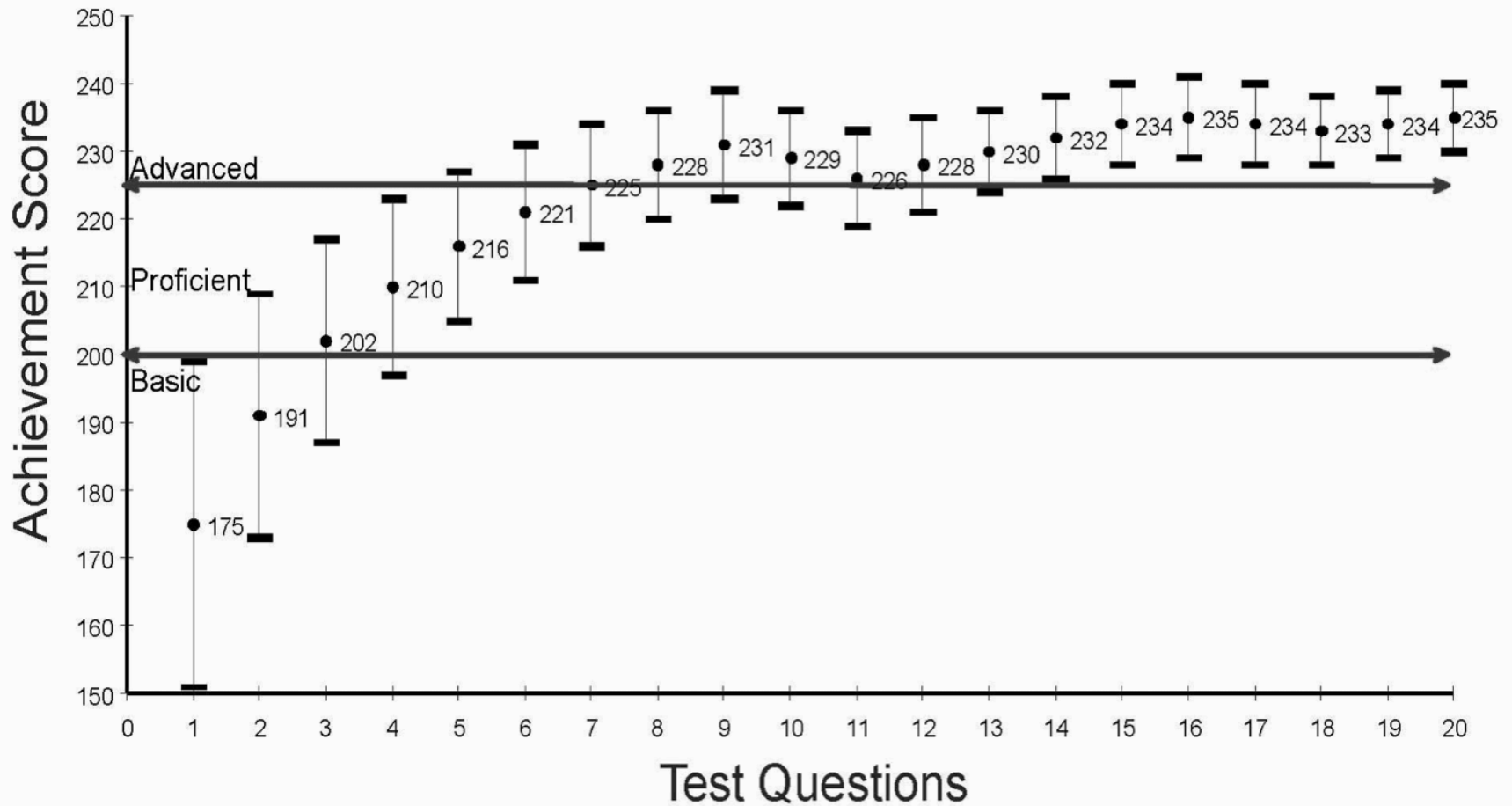State Superintendent
of Public Instruction

TOM TORLAKSON
State Superintendent
of Public Instruction

CAT Testing Process

Develop item bank and CAT system

Specify starting point and start test

Select item and deliver to examinee

Score item, update examinee score estimate

Evaluate termination criterion

Not satisfied

Satisfied

Conclude test

From: Framework for the Development of Computerized Adaptive Tests,
Thompson, Nathan A. and Weiss, David J. January 2011

10

Computer Adaptive Testing
20 Questions

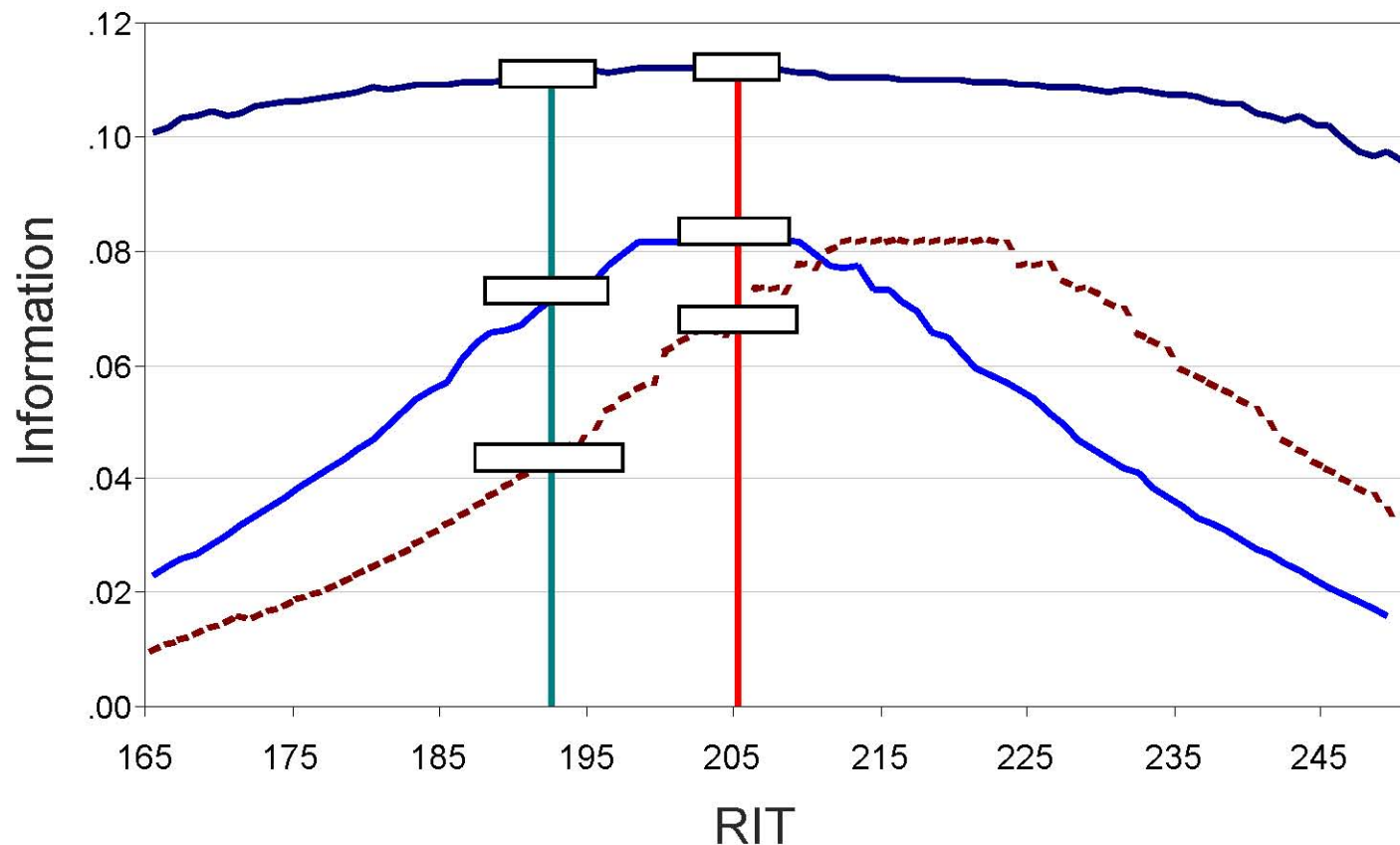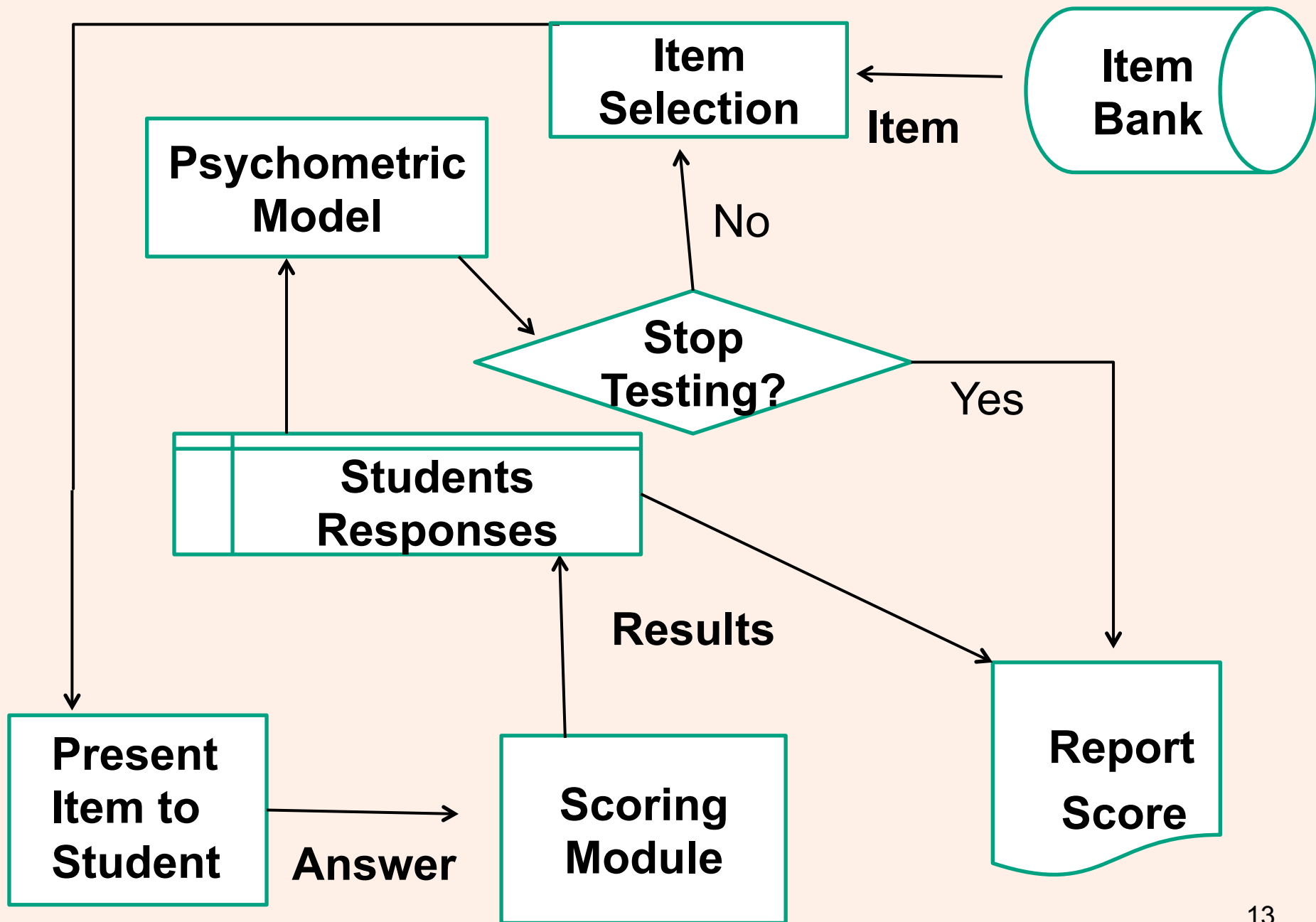# Test Information Functions for CAT and P&P Tests

Students' Mean = **211.7**　　　s.d. = **11.11**　　　Proficiency = **205**　　　Basic = **192**

Test Information Functions for Grade 4 Mathematics

Item Selection

Item Bank

Psychometric Model

Item

Stop Testing?

No

Yes

Students Responses

Results

Present Item to Student

Answer

Scoring Module

Report Score

13

Important Issues in CAT

- Requires large number of extremely high quality items.

- Different examinees take different forms of the test and possibly different numbers of items

- Score is not based on how many items student answers correctly, but rather which items

- Item selection based on item parameters (difficulty) and content specifications

14

Important Issues in CAT (cont.)

- Starting and stopping points must be defined in terms of test information function/ error of measurement

- Need to equate P&P tests with CBT tests during transition period

Scoring Technology

- Traditional machine scoring
  - Dichotomous (correct/incorrect) scoring most common
  - Exact word, number, or grid matches
  - No partial credit

- Automated scoring
  - Allows scoring of short answer and essay questions
  - Require set of human scored papers to develop the scoring model
  - Can give partial credit, or multiple point scores

TOM TORLAKSON
State Superintendent
of Public Instruction

How Automated Scoring Works

- Uses a set of human scored examples to develop a statistical model used to analyze answers (e.g., latent semantic analysis or natural language processing)

- Generally examine overall form and specific combinations of words

- Has an extensive library of possible meanings for words

What can be scored?

- Written responses
  - Prompt specific essays
  - Prompt independent essays
  - Short answers
  - Summaries

- Spoken language
  - Correctness
  - Fluency

- Responses to simulations
  - Diagnosis of a patient's illness
 - Landing a plane

TOM TORLAKSON
State Superintendent
of Public Instruction

How good is automated scoring?

- Automated scores are consistent with the scores from expert human graders

- The way automated scores are produced is understandable and substantively meaningful

- Automated scores are fair

- Automated scores have been validated against external measures in the same way as is done with human scoring

- The impact of automated scoring on reported scores is understood

Source: ETS, Pearson, and the College Board's recent report: "*Automated Scoring for the Common Core Standards*."
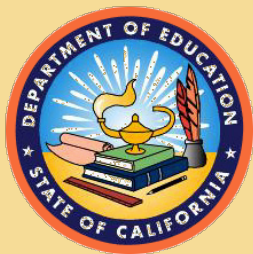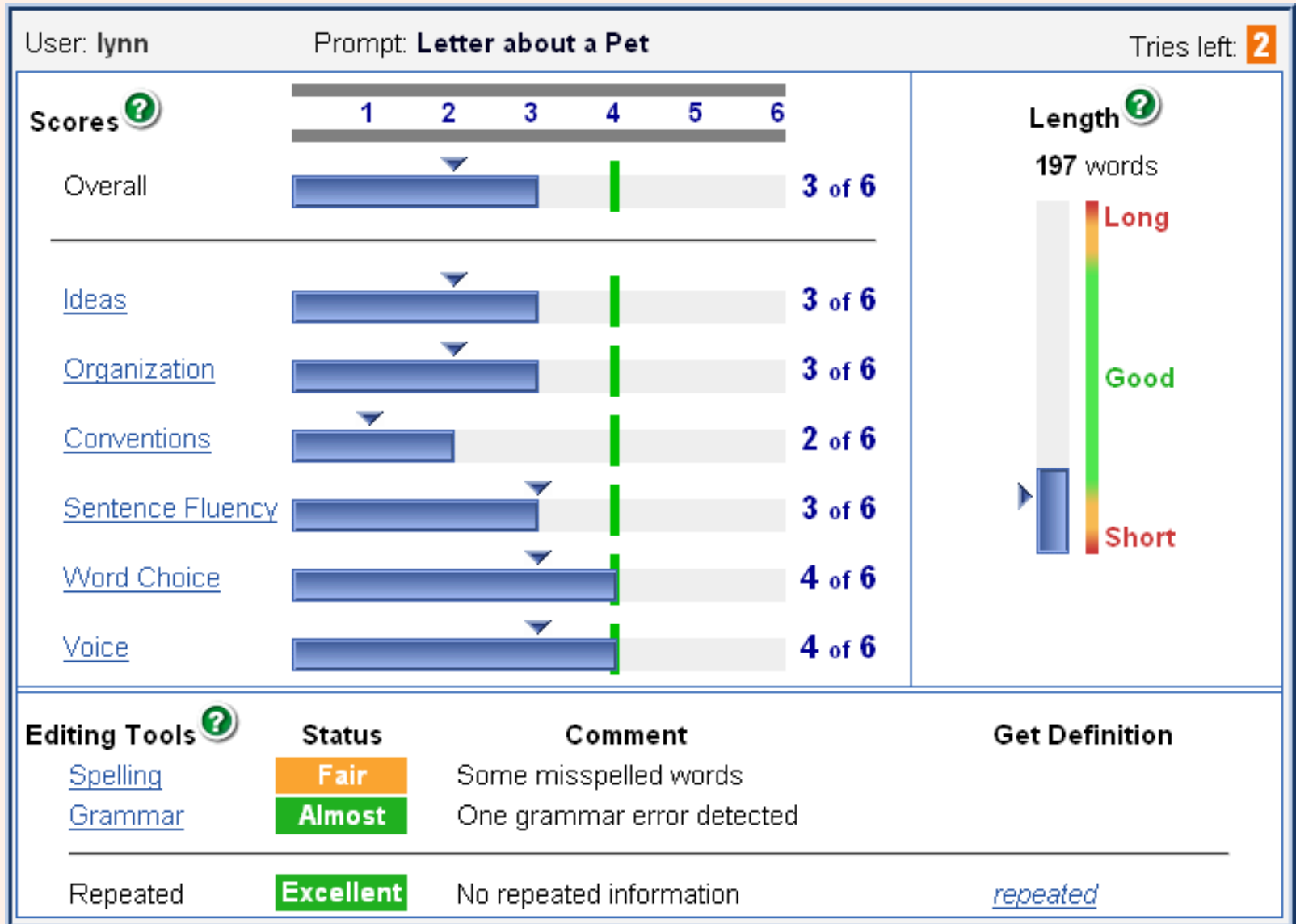
Autoscoring Performance

Response	Assessment Prompt Material	N	Machine–Human Correlation	Human–Human Correlation	Source
Written	81 published essay prompts (grade 6-12)	400	0.89	0.86	Prentice Hall
	18 research-leveled essay prompts (grade 4-12)	635	0.91	0.91	MetaMetrics
	5 synthesizing memos from multiple sources	1239	0.88	0.79	Council for Aid to Education
Spoken	2000 spoken English items	50	0.97	0.98	Balogh & et al. (2005)
	3000 spoken Arabic items	134	0.97	0.99	Bernstein et al. (2009)
	9 Oral Reading Fluency Passage Grades 1-5	248	0.98	0.99	Downey et al. (2011)

Source: Streeter et. al. *Pearson's Automated Scoring of Writing, Speaking, and Mathematics*, Pearson, May 2011

TOM TORLAKSON
State Superintendent
of Public Instruction

Example Essay Feedback

| User: **lynn** | Prompt: **Letter about a Pet** | Tries left: **2** |

Scores

	1	2	3	4	5	6	
Overall							3 of 6
Ideas							3 of 6
Organization							3 of 6
Conventions							2 of 6
Sentence Fluency							3 of 6
Word Choice							4 of 6
Voice							4 of 6

Length

197 words

Long — Good — Short

Editing Tools	Status	Comment	Get Definition
Spelling	Fair	Some misspelled words	
Grammar	Almost	One grammar error detected	
Repeated	Excellent	No repeated information	*repeated*

Source: Streeter et. al. *Pearson's Automated Scoring of Writing, Speaking, and Mathematics*, Pearson May 2011

Data Requirements for Automated Scoring of Various Item Types

Item Type	Response Length in Words	Typical Data Requirements for development	Measures Returned
Prompt-Specific Essays	100-500	200-250 double-scored student essays	Overall score, trait scores, grammar & mechanics feedback
Prompt Independent Essays (general models)	100-500	Approximately 1000 essays per grade	Overall score, select trait scores, grammar & mechanics feedback
Short Answers	~10-60	500 double-scored student answers	Total or partial-credit content score
Summaries	50-250	Readings to be summarized divided by major sections	Content coverage score for each section; checks copying, length, redundancy and irrelevance.

Questions?

Update on Key Consortium Activities

- Executive Committee Co-Chair Election

- Preliminary Test Blueprints

- Revised Draft Initial Achievement Level Descriptors

- Upcoming Opportunity for Teacher Involvement

- Spring 2013 Pilot Test

- Technology Update

TOM TORLAKSON
State Superintendent
of Public Instruction

Executive Committee
Co-Chair Election

- Last week, Smarter Balanced governing states elected Deb Sigman, CDE Deputy Superintendent, as Executive Committee Co-Chair

- Deb officially took office as Co-Chair on February 12, 2013

TOM TORLAKSON
State Superintendent
of Public Instruction

Preliminary Test Blueprints

- Approved by governing states in November 2012

- Include critical information about the number of items, score points, and depth of knowledge for items associated with each assessment target

- Guide the development of items and performance tasks, the pilot and field tests, score reporting, standard setting, and ongoing research

- Are considered preliminary until after review of the data gathered from the pilot and field tests

- Links to blueprints available under the heading "Preliminary Test Blueprints" on the Smarter Balanced Web page at http://www.smarterbalanced.org/smarter-balanced-assessments/

TOM TORLAKSON
State Superintendent
of Public Instruction

Revised Draft Initial Achievement Level Descriptors (ALDs)

TOM TORLAKSON
State Superintendent
of Public Instruction

- Initial draft ALDs were released for public comment November 27, 2012, through January 15, 2013.

- A second window is open February 4 through 20 to provide public comments on the revised ALDs.

- Feedback must be submitted via the online survey provided by Smarter Balanced.

- The revised draft ALD documents, online survey for providing feedback, and recording of the February 6 Webinar highlighting the revisions are available at http://www.smarterbalanced.org/achievementlevel-descriptors-and-college-readiness.

- Governing states are expected to adopt the initial ALDs in Spring 2013.

Upcoming Opportunity for Teacher Involvement

- Smarter Balanced digital library of formative assessment tools and practices

- State Network of Educators to be formed to review proposed tools and practices for inclusion in the digital library

- Recruitment of State Network of Educators expected to begin Spring 2013

Spring 2013 Pilot Test – Scientific Sample

TOM TORLAKSON
State Superintendent
of Public Instruction

- Test window has been extended: February 20–May 24, 2013.

- CDE encourages all selected schools to participate (January 29 letter from CDE).

- CDE is assisting participating schools by submitting required student data (February 11 letter from CDE).

- Search for selected CA schools by county and district and view participation status on the CDE Smarter Balanced Web page at http://www.cde.ca.gov/ta/tg/sa/smarterbalanced.asp. Select the "Spring 2013 Pilot Test" hyperlink.

- More than 1,000 CA schools have confirmed participation.

- **Deadline for schools to confirm participation has been extended to February 28.**

Spring 2013 Pilot Test – Volunteer Sample

TOM TORLAKSON
State Superintendent
of Public Instruction

- Open to all schools in member states

- Testing window: April 9 through May 10, 2013

- Participation in volunteer pilot available any time during the testing window

- Schools volunteer by completing the volunteer survey at: https://www.surveymonkey.com/s/SmarterBalancedPilot

- The deadline for registering to participate is March 27, 2013.

Smarter Balanced Pilot Test Contacts and Information

Recruitment Information:

Data Recognition Corporation

800-847-3193

smarterbalancedrecruitment@datarecognitioncorp.com

All Other Information:

American Institutes for Research

866-815-7246

smarterbalancedhelpdesk@air.org

SBAC Pilot Test Frequently Asked Questions:

http://www.cde.ca.gov/ta/tg/sa/smarterbalanced.asp

Select the "Spring 2013 Pilot Test" hyperlink.

Technology Update

- iPads will not be supported for schools assigned to the February 20–March 6 pilot test window. If schools intend to use iPads, the school will be assigned to a later window.

- Text-to-speech on iPads will not be available for the pilot test.

- External keyboards will be required for tablet devices.

32

For Further Information

CDE Transition Office
sbac@cde.ca.gov
916-445-8517

Technology Readiness Coordinator
sbac-itreadiness@cde.ca.gov

**Smarter Balanced
Assessment Consortium Web Site**
http://www.smarterbalanced.org/

CDE Smarter Balanced Web Page http://www.cde.ca.gov/sbac/